Ecological functions of uncultured microorganisms in the cobalt-rich ferromanganese crust of a seamount in the central Pacific are elucidated by fosmid sequencing

HUO Yingyi[1,2], CHENG Hong[3], Anton F. POST[4], WANG Chunsheng[1,2], JIANG Xiawei[5], PAN Jie[3], WU Min[3*], XU Xuewei[1,2*]

[1] Laboratory of Marine Ecosystem and Biogeochemistry, Second Institute of Oceanography, State Oceanic Administration, Hangzhou 310012, P. R. China

[2] State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, Hangzhou 310012, P. R. China

[3] College of Life Sciences, Zhejiang University, Hangzhou 310058, P. R. China

[4] The Josephine Bay Paul Center for Comparative Molecular Biology and Evolution, Marine Biology Laboratory, Woods Hole, MA 02543, USA

[5] State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, the First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310003, P. R. China

Foundation item: China Ocean Mineral Resources R & D Association (COMRA) Special Foundation (No. DY125-15-R-03 and DY125-13-E-01); the National Natural Science Foundation of China (No. 41276173); the Zhejiang Provincial Natural Science Foundation of China (No. LQ13D060002) and the Scientific Research Fund of the Second Institute of Oceanography, SOA (No. JT1305).

Corresponding authors: XU Xuewei, Email: xuxw@sio.org.cn, Tel: +86-571-81963208, Fax: +86-571-88071539; Wu Min, E-mail: wumin@zju.edu.cn, Tel: +86-571-88206261, Fax: +86-571-88206261.

Summary

Cobalt-rich ferromanganese is an important seafloor mineral and is abundantly present in the seamount crusts. Such crusts form potential hotspots for biogeochemical activity and microbial diversity, yet our understanding of their microbial communities is lacking. In this study, we used a cultivation-independent approach to recover genomic information and derive ecological functions of the microbes in a sediment sample collected from the cobalt-rich ferromanganese crust of a seamount region in the central Pacific. A total of 78 distinct clones were obtained by fosmid library screening with a 16S rRNA based PCR method. Proteobacteria and MGI Thaumarchaeota dominated the bacterial and archaeal 16S rRNA gene sequence results in the microbial community. Nine fosmid clones were sequenced and annotated. Numerous genes encoding proteins involved in metabolic functions and heavy metal resistance were identified, suggesting alternative metabolic pathways and stress responses that are essential for microbial survival in the cobalt-rich ferromanganese crust. In addition, genes that participate in the synthesis of organic acids and exoploymers were discovered. Reconstruction of the metabolic pathways revealed that the nitrogen cycle is an important biogeochemical process in the cobalt-rich ferromanganese crust. In addition, horizontal gene transfer (HGT) events have been observed, and most of them came from bacteria, with some occurring in archaea and plants. Clone W4-93a, belonging to MGI Thaumarchaeota, contained a region of gene synteny. Comparative analyses suggested that a high frequency of HGT events as well as genomic divergence play important roles in the microbial adaption to the deep-sea environment.

Key words: seamount; cobalt-rich ferromanganese crust; metagenome; horizontal gene transfer

1 Introduction

Seamounts are widespread and defined as topographic rises from the ocean floor with a limited area across the summit, which is below sea level or emerges above the sea surface only for short periods of time (Menard, 1964; Staudigel et al., 2010). Recently, 33,452 seamounts (elevation of > 1 000 m) were identified in global bathymetric datasets, most of them (57.2%) located in the Pacific Ocean (Yesson et al., 2011). However, the number of seamounts remains under debate due to the different definitions of what constitutes a seamount as well as the variation in techniques used to count them (Hillier and Watts, 2007; Iyer et al., 2012; Wessel et al., 2010). Although seamounts have a high degree of biodiversity, harbour unique biological communities, display high levels of endemism, represent hotspots of nutrient cycling and support commercial fisheries, fewer than 300 seamounts have been thoroughly sampled, and the majority of these studies have focused on hydrothermal vents (Clark et al., 2010; Duffy, 2008; Emerson and Moyer, 2010; Rowden et al., 2010; Schlacher et al., 2010).

Deposits of cobalt-rich, oxidised ferromanganese in crusts that cover seamounts were first discovered in 1980 (Craig et al., 1982; Ito et al., 2008; Muiños et al., 2013). These crusts usually grow at very slow rates (1-10 μm per 10^3 years) and exist only on a few seamounts (Fu et al., 2005; Koschinsky and Hein, 2003). Cobalt-rich ferromanganese crusts are a rich source of metals such as ferromanganese oxide, cobalt, copper, nickel, platinum and other rare earth elements (Fu et al., 2005; Fuyuan et al., 2008). Currently, our knowledge of biological interactions at the cobalt-rich ferromanganese crusts is very limited, and more research is needed into their resident microbial communities and their ecosystem functions to evaluate the environmental impacts of future crust exploration and mining.

The role of biogenesis in cobalt-rich ferromanganese crust formation on seamounts remains controversial and poorly understood. Previous results from a northern Pacific seamount indicated that crust accretion is not a purely physicochemical process as it also involves microbial processes (Verlaan, 1992). Microorganisms act as biological nuclei for the formation of cobalt-rich crusts, suggesting that biomineralization is indispensable in the mineral formation process (Wang and Müller, 2009). Scanning electron microscopy studies suggest that biological processes are involved in the formation of the ferromanganese crusts covering seamounts in the central Atlantic (Wang et al., 2011). Several studies of microbial diversity in sediments from cobalt-rich ferromanganese crusts of central Pacific seamounts have been performed using culture-independent approaches (Jiang et al., 2012; Liao et al., 2011). Phylogenetic analyses of bacterial and archaeal 16S rRNA clone libraries have

revealed the predominance of *Proteobacteria* and marine archaeal group I (MGI), and it has been suggested that members of this community may be involved in sulphur, nitrogen and metal cycling in cobalt-rich ferromanganese crusts (Liao et al., 2011). Recently, nine novel lipolytic enzymes were identified, suggesting that the microbial populations participate in carbon degradation, calcium deposition and contribute to biomineralization (Jiang et al., 2012).

Recent work revealed the microbial community structure and diversity of the microbes in cobalt-rich ferromanganese crusts by using molecular and electron microscopy approaches. However, the functional aspects of the microbes should not be overlooked. In this study, a metagenomic library of deep-sea sediment collected from a cobalt-rich ferromanganese crust region from a seamount was screened to describe the genome content and biological properties of uncultivated microorganisms. In total, approximately 21,000 randomly selected fosmid clones were subjected to PCR-based screening, 35 archaeal and 43 bacterial 16S rRNA gene-containing clones were obtained, and nine fragments were sequenced and analysed. To our knowledge, this is the first report of a metagenomic approach to identify the role of microorganisms in the formation of cobalt-rich ferromanganese crusts of seamounts.

2 Materials and Methods

2.1 *Sample collection, geochemical properties analysis and library construction*

A deep-sea sediment sample was collected from the skirt of a seamount located in the cobalt-rich crust deposit region in the central Pacific Ocean. The sample collection method, location, DNA extraction procedure and library construction information were described in detail in an earlier publication (Jiang et al., 2012). The elemental composition of the sample was determined by the Zhejiang Institute of Geology and Mineral Resources using various methods, including the gravimetric method (for SiO_2), inductively coupled plasma-atomic emission spectrometry (ICP-AES; for Al_2O_3, Fe_2O_3, CaO, MgO, TiO_2, MnO and P_2O_5), atomic absorption spectrometry (AAS; for K_2O and Na_2O) and inductively coupled plasma-mass spectrometry (ICP-MS; for Ni, Co, V, Cr, Cu, Zn, Cd, Pb, Mo and Ba).

2.2 *16S rRNA gene screening*

Approximately 21,000 randomly selected fosmid clones were subjected to PCR-based screening. The DNA of pooled fosmid clones was extracted using the Axygen Plasmid Miniprep Kit (Axygen Biotechnology, Hangzhou, China). The fosmid DNA templates were treated with the plasmid-safe ATP-dependent DNase (Epicentre Biotechnologies, Madison,

Wisconsin, USA) to remove the chromosomal DNA contamination of the host strain (*Escherichia coli* EPI300). Primers Ar20F (5'-TTCCGGTTGATCCYGCCTGA-3') and Arch958R (5'-TCCGGCGTTGAMTCCAATT-3') (DeLong, 1992) were used for archaeal 16S rRNA gene amplification. Primers 27F (5'-AGAGTTTGATCCTGGCTCAG-3') and 23S1R (5'-GGGTTTCCCCATTCGGAAATC-3') were used for the identification of the bacterial 16S rRNA gene with the adjacent intergenic spacer region (ISR) (Garcia-Martinez et al., 1996). Thirty-five cycles of amplification were carried out under the following conditions for the archaeal 16S rRNA gene: denaturation at 94°C for 45 seconds, annealing at 55°C for 45 seconds and elongation at 72°C for 1 minute. Thirty-five cycles of amplification of bacterial 16S rRNA gene were performed under the following conditions: denaturation at 94°C for 15 seconds, annealing at 50°C for 30 seconds and elongation at 72°C for 2 minutes (Martín-Cuadrado et al., 2007). PCR fragments were extracted from the gel using the AXYGEN gel extraction kit (AXYGEN, Hangzhou, China) and then cloned into the pMD19-T vector (TAKARA, Dalian, China). The 16S rRNA gene fragments were sequenced with the primers M13F/M13R (archaea) and 27F/1492R (bacteria). The closest relatives of the 16S rRNA sequences were obtained from the NCBI GenBank database using blastn. Evolutionary distances were calculated according to Kimura's two-parameter correction method. Neighbour-joining trees were constructed with a bootstrap value of 500 using MEGA version 5.0 (Tamura et al., 2011). The accession numbers of the 16S rRNA gene sequences are JQ013299-JQ013333 (archaeal) and JQ013334-JQ013376 (bacterial).

2.3 *Fragments sequencing and analysis*

Two archaeal and seven bacterial fosmid clones were selected and their sequences were determined by sequencing using a Roche 454 GS-FLX and Illumina/Solexa Genome Analyzer II platforms (Tongji-SCBIT Biotechnology Co., Ltd). Gaps were closed with the help of targeted PCR. The PCR products were sequenced by primer walking. Open reading frames (ORFs) were predicted using MetaGeneMark (Zhu et al., 2010). Gene identification was obtained by submitting the deduced protein sequences for ortholog/homolog searches in the NCBI nr database, the Cluster of Orthologous Groups (COG) database (Tatusov et al., 2000) and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa and Goto, 2000) using blastp. A threshold e-value of 1e-5 was used for all analyses. For phylogenetic analysis of ORFs, a blastp search of protein sequences in the NCBI nr database was carried out with default parameters, and the sequences for the best blast hits were retrieved from the database. Neighbour-joining trees were constructed with a bootstrap value of 500 using the

Poisson model option in the MEGA 5.0 phylogenetic software package (Tamura et al., 2011). The Genbank accessions of the fosmid sequences are JQ085817-JQ085825.

3 Results

3.1 *Elemental composition*

The elemental composition of the sediment sample SEAM02 is shown in Table 1. Compared with offshore sediment (Zhao, 1988) and continental crust (Wedepohl, 1995), the concentrations of Al_2O_3 and CaO were much lower in the sample, whereas those of Na_2O, MnO, Ni, Co, Cu, Zn, Pb and Ba were higher. Specifically, Na_2O, MnO, Co, Cu and Ba, reached concentrations of more than twice those found in the offshore sediment and continental crust. Taken together, the sample SEAM02 was mainly composed of SiO_2 (54.07%), Al_2O_3 (8.35%), Na_2O (7.18%) and Fe_2O_3 (4.14%) and was rich in metals, including Mn, Ni, Co, Cu, Zn, Pb and Ba.

3.2 *Microbial community composition*

A total of 35 archaeal 16S rRNA-containing clones were obtained from the fosmid library. All of the archaeal clones belonged to Marine Group I (MGI) in the phylum Thaumarchaeota except clone W5-61a, which was grouped into Marine Benthic Group A (MBGA) (Fig. 1a). These sequences were most closely related to those from other deep-sea sediment environments, including sediments from the Pacific nodule province (Xu et al., 2005), the Weddell Sea of Antarctica (Gillan and Danis, 2007), the southern Mariana Trough (Kato et al., 2009a; Kato et al., 2009b), the east Pacific rise (Ehrhardt et al., 2007; Li et al., 2008), the Mediterranean cold seep (Heijs et al., 2007) and the Barents Sea cold seep (Losekann et al., 2007). This result indicated that our clones represent members of a common and abundant archaeal community in deep-sea sediments. In accordance with findings from previous microbial diversity studies of deep-sea sediment with cobalt-rich crust deposits (Liao et al., 2011), MGI was the dominant archaeal group. The sequence of clone W5-61a, the only clone that affiliated with members of group MBGA, shared <90% identity with previously reported clones, with the exception of clones aEPR13S208 (97.6%) and YS16As04 (92.4%) retrieved from the East Pacific Rise (Li et al., 2008) and the Southern Mariana Trough (Kato et al., 2009a), respectively. There are no cultivated species reported for MBGA members, and only a single strain, *Nitrosopumilus maritimus* SCM1, within the MGI group has been isolated so far (Könneke et al., 2005), leaving us with little insight into the genetic make-up of their genomes and the physiological functions they encode.

A total of 43 bacterial fosmid clones were obtained that distributed over 7 bacterial groups (Fig. 1b): Alphaproteobacteria (6 clones), Betaproteobacteria (1 clone), Gammaproteobacteria (14 clones), Deltaproteobacteria (3 clones), Actinobacteria (11 clones), Gemmatimonadetes (3 clones) and Chlorobi (1 clone). The other 4 clones (W4-21b, W4-50b, W5-15b and W5-77b) could not be assigned to any taxonomic division. A total of 38 bacterial clones showed high identities with uncultured clones from the deep-sea surface sediments of the south Atlantic Ocean (Schauer et al., 2009), the seafloor lavas of the east Pacific rise (EPR) and the Hawaiian basalts (Santelli et al., 2008), indicating these bacteria may be common to these deep-sea environments. In addition, the 16S rRNA gene sequences of clones W4-21b, W5-15b and W5-102b did not have matches with >90% identity to other sequences in the database, suggesting they might represent novel taxa for the deep-sea environment, and they may have unique adaptation to the cobalt-rich crust environment.

3.3 *Gene content of fosmid clones*

To obtain more genomic information on microbial adaptation to deep-sea sediments, we sequenced two archaeal and seven bacterial genome fragments. The fosmid insert sizes ranged from 23 to 45 kb, with G+C content ranging from 36% to 65%. A detailed description of DNA insert sizes, ORF positions, predicted functions, closest relatives, and COG classification are summarised in Table S1. In addition, inserts were subjected to gene annotation, revealing a range of gene functions.

Archaeal fosmid clones

To the best of our knowledge, clone W5-61a is the first genome fragment to be sequenced for a MBGA member of the Thaumarchaeota. The fragment has a G+C content of 48.6%, considerably higher than that of MGI Thaumarchaeota (approx. 34%) (Gilbert et al., 2011; Walker et al., 2010). Clone W5-61a (32,142 bp) contains 25 ORFs, of which 23 of them could be assigned with the COG classification system. Of these, 17 encode functions in basic metabolic processes: leucine biosynthesis (ORFs 1-3), ion transport (ORFs 7-9), DNA protection and repair (ORFs 13-14) and purine metabolism (ORFs 15-22). Most of the predicted genes had the most significant blast hits to members of archaea, whereas ORF2, ORF3 and ORF6 showed the highest identity to bacterial genes. ORF2 and ORF3 encoded the large and small subunits of 3-isopropylmalate dehydratase, and phylogeny indicates that they may have been transferred from Firmicutes via HGT (Fig. S1a,b). Both ORF6 and ORF11 were identified as encoding the pyrroloquinoline quinone biosynthesis protein C (PqqC);

however, only 53.4% of the amino acid residues were similar between them. The phylogenetic tree was constructed based on bacterial and available archaeal original PqqC genes (Fig. S1c). The two genes formed distinct phyletic lines with high bootstrap values towards the periphery of the bacterial lineage, revealing that they may be of bacterial origin. ORF12 encodes a PQQ-dependent alcohol dehydrogenase, which is rarely found in archaea. The function of archaeal PQQ-dependent alcohol dehydrogenase is still unknown; however, bacterial alcohol dehydrogenase had been proved to oxidize various alcohols for bioenergy generation (Adachi et al., 2007). This gene might be important for the energy requirement of Archaea in deep sea sediments.

Clone W4-93a (34,190 bp) was most closely related to *Candidatus* Nitrosoarchaeum limnia within the marine group I Thaumarchaeota (> 95.8% 16S sequence identity). The G + C content of this fragment was 36.3%, and a total of 49 ORFs were predicted. Deduced amino sequences indicated that only 21 proteins (44.7%) could be assigned a physiological function with the COG classification system, and 9 proteins (19.1%) did not show significant similarity to any proteins in the NCBI nr database. Apart from these 9 predicted proteins that had no significant relatives, the other ORFs were most closely related to known archaeal genes found in the members of Thaumarchaeota (17 ORFs to *Candidatus* Nitrosoarchaeum limnia, 16 ORFs to *Nitrosopumilus maritimus*, 4 ORFs to *Candidatus* Cenarchaeum symbiosum and 2 ORFs to *Candidatus* Nitrosoarchaeum koreensis). Interestingly, ORF25 was most closely related to the nitrogen regulatory protein P-II from *Thermococcus sibiricus*, a hyperthermophilic member of the Euryarchaeota (Miroshnichenko et al., 2001) (Table S1). Phylogenetic analysis also indicated that this gene may have been obtained from euryarchaeotal species by HGT (Fig. S1d).

Gene organisation and synteny was determined by comparing clone W4-93a with two MGI Thaumarchaeota fragments (Fig. 2), *Nitrosopumilus maritimus* SCM1 and clone 74A4 (Béjà et al., 2002). *N. maritimus* SCM1 isolated from seawater from the Seattle Aquarium was the first member of the Thaumarchaeota to have been cultured (Könneke et al., 2005), and this isolate was identified as a key chemolithoautotrophic ammonia-oxidiser in the marine environment. Fosmid clone 74A4 was obtained from a surface sample in the Southern Ocean (Béjà et al., 2002). The 16S rRNA gene sequence of clone W4-93a was > 95% identical to that of *N. maritimus* SCM1 as well as clone 74A4. Comparative genomic analysis showed that the gene content and arrangement in the three related MGI Thaumarchaeota genome fragments were largely conserved at the 5'-end of the 16S-23S rRNA operon, but not at the 3'-end (Fig. 2). Upstream of the 16S-23S rRNA operon, a colinear region spanning 12 genes

(approx. 13 kbp), was shared by the three fragments. Downstream of the 16S-23S rRNA genes, however, strain SCM1 harboured an approximately 26.5 kbp fragment, whereas both clone W4-93a and 74A4 did not contain this region. A single gene (ORF48, biotin--acetyl-CoA-carboxylase ligase) was common to the three fragments. Although clone W4-93a lacked a genome fragment near the 16S-23S rRNA genes, it contained a nitrogen regulatory protein P-II gene (ORF25) and a nitroreductase gene (ORF31) that participates in nitrogen cycling.

Bacterial fosmid clones

Clone W4-39b contained 28 ORFs and a 16S-23S-5S rRNA operon with two tRNA genes (tRNA-Ala and tRNA-Ile). The 16S rRNA and 23S rRNA sequences identified this clone as belonging to the Deltaproteobacteria. COG classification was successful in assigning functions to most of the predicted proteins, including genes responsible for L-glutamate synthesis (ORF7-8), glycerophospholipids metabolism (ORF10-11), *de novo* purine biosynthesis (ORF18), substrate transport (ORF13, ORF15-16 and ORF23) and transposition (ORF20-22). ORF20-22 encoded transposases that clustered with Alphaproteobacteria in a monophyletic tree (Fig. S1f). ORF24-26 had a small number of close relatives, including the genera *Erythrobacter*, *Novosphingobium*, *Sphingobium* and *Sphingomonas* of Alphaproteobacteria and the genera *Gallionella*, *Methylotenera* and *Limnobacter* of Betaproteobacteria (Fig. S1g-i). Intracellular ammonium is incorporated into carbon skeletons via the glutamate / glutamine synthase pathway. Interestingly, ORF7 and ORF8, which encoded small and large subunits of glutamate synthase, were most similar to those found in Actinobacteria (Fig. S1e).

Clone W4-87b exhibited the highest 16S and 23S rRNA gene sequence identities to *Nitrosospira multiformis* ATCC 25196 (97.8% and 97.1%, respectively), which is an ammonia-oxidising bacterium (AOB) isolated from soil (Norton et al., 2008). Among the predicted 18 ORFs, only seven ORFs were related to the Betaproteobacteria class. ORF4 showed a clearly Bacteroidetes origin (Fig. S1j). Clone W4-87b encoded a urease accessory protein (UreD) involved in the activation of a urease that hydrolyses urea to ammonia. Considering the phylogenetic analyses of the rRNA genes, as well as that of *ureD*, we considered that clone W4-87b might have derived from an AOB similar to *N. multiformis* which plays an important role in the nitrogen cycle. In addition, some oxidative stress resistance genes, including thioredoxin (ORF3 and ORF13), superoxide dismutase (ORF9) and universal stress protein A (ORF10) were found in clone W4-87b.

A similar situation was observed for clone W5-47b, which showed the highest 16S rRNA

sequence identity to *Ignavibacterium album* Mat9-16 (92.6%) belonging to the Chlorobi. Chlorobi are obligate anaerobic photoautotrophic bacteria. Clone W5-47b contained several antibiotic-synthesising and metabolic genes from other bacteria, as well as archaea, to survive in the deep-sea environment. Examples of these were ORF1 (penicillin synthesis), ORF2-5 (pentose phosphate pathway), ORF10 (streptomycin biosynthesis), ORF17 and ORF24 (purine and pyridine metabolism) and ORF22 (poly-γ-glutamate synthesis).

The 16S and 23 rRNA genes found in clone W5-102b identified this fragment as being derived from a member of the Actinobacteria. Most ORFs (54.8%) were predicted to be hypothetical proteins, and 13 ORFs had no orthologs in the NCBI nr database. Some ORFs might have been acquired by HGT, including ORF6 from Firmicutes (Fig S1k), ORF10 from other bacteria (Fig. S1l), ORF18 from Euryarchaeota (Fig. S1m), ORF25 and ORF30 from Chloroflexi and ORF26 from Bacteroidia.

Tree topology with a high bootstrap value (100%) revealed that clone W5-51b fell within a cluster composed of Gemmatimonadetes members, and formed an independent clade (Fig. 1b). However, this clone was distinguishable from the known Gemmatimonadetes species based on the low (< 85%) (Table S1) identities of its 16S rRNA gene and other functional genes. Upstream of the 5S-23S-16S rRNA genes, clone W5-51b fragments were partially conserved, and most ORFs (10/13, 76.9%) exhibited the highest sequence identities with genes found in *Gemmatimonas aurantiaca*. Downstream of the rRNA gene cluster, ORFs were similar to those of other Gemmatimonadetes members, and most of them were more closely related to Proteobacteria. Phylogenetic analysis revealed that ORF17 and ORF18-19 might have been acquired by HGT from Alphaproteobacteria and Gammaproteobacteria, respectively (Fig. S1n-o). Our data suggest that clone W5-15b presented a new lineage in the phylum Gemmatimonadetes.

Clones W4-21b and W5-15b belonged to unknown taxonomic groups within the phylogenetic tree (Fig. 1) and showed very low 16S rRNA sequence identity with known bacterial species (< 78%, Table S1). Some important functional genes were detected in these two clones. In clone W4-21b, ORF22 encoded a SpoIID/LytB domain-containing protein that is involved in sporulation (Lopez-Diaz et al., 1986); ORF24 encodes a heavy metal resistance protein (CzcD) that is able to mediate metal efflux and so enhance the cell's resistance to e.g., cobalt, zinc and cadmium (Nies, 1992); and ORF15 (RadC), ORF16 (RecJ), ORF29 (DNA polymerase) and ORF30 (LexA repressor) were related to DNA repair. Sporulation, heavy metal resistance and DNA repair are important mechanisms of environmental stress resistance in microbes, indicating that clone W4-21b belonged to a spore-forming bacterium adapted the

deep-sea cobalt-rich ferromanganese crust environment. In clone W5-15b, ORF1 (adenylate/guanylate cyclase), ORF3 (CspE), ORF6 (PspC), ORF9 (PhoH), ORF22 (glutathione S-transferase) and ORF26 (RecB) were recognized as the stress response regulating genes, and ORF32 encoded a poly-γ-glutamate synthesis protein. Poly-γ-glutamate is a natural polymer synthesised by gram-positive bacteria. It allows bacteria to survive at high salt concentrations and may also act as a virulence factor or a storage element for carbon and nitrogen precursors or as an energy source (Candela and Fouet, 2006). The above genes reveal that clone W5-15b was derived from a gram-positive bacterium and was able to resist various stresses, including low temperature, high osmotic pressure and low nutrient availability. In addition, potential HGT of ORF22 from plant was identified (Fig. S1q).

4 Discussion

Recently, we published the first assessment of the bacterial and archaeal diversity in the sediment collected from a cobalt-rich ferromanganese crust (Liao et al., 2011). Proteobacteria and MGI Thaumarchaeota dominated the bacterial and archaeal communities, respectively. In addition, the microbial diversity inside nodules and in the surrounding sediments collected from a cobalt-rich ferromanganese crust were compared (Wu et al., 2013). Here, we focused on the ecological functions of these microbial communities with a special emphasis on their adaptive ability and survival in cobalt-rich ferromanganese crusts.

Manganese, cobalt, copper and nickel are the most important metal elements available in cobalt-rich ferromanganese crust. The content of Co + Cu + Ni is one of the significant evaluation indicators of the crust. Our study revealed that the concentrations of Mn, Ni, Co and Cu in the sediment from station SEAM02 were more than twice those found in offshore sediments or in continental crust (Table 1), whereas they were at much lower concentrations than those in the cobalt-rich ferromanganese crust (data not shown). Some oxidative stress resistance genes, including thioredoxin, superoxide dismutase and PqqC, were detected in our fosmid library. A gene encoding a cobalt-zinc-cadmium resistance protein (ORF24 in clone W4-21b) was also found, suggesting that some microbes that inhabit in the sediment adapt to heavy metal toxicity (Supplementary Table 1). Recent works demonstrate that cobalt-rich ferromanganese crusts are formed by biologically driven processes involving microbes (Wang et al., 2009; Wang and Müller, 2009), and that microorganisms are responsible for the bulk of Mn oxide formation (Tebo et al., 2004). Some microorganisms produce high amounts of organic acids or exopolymers to aggregate the metal granules (Gadd, 2007; Guibaud et al., 2009), and the surfaces of the budding and sheathed bacteria are surrounded by Mn oxide,

346 which forms a shell and protects the microbes from heavy metal invasion (Ghiorse, 1984;
347 Santelli et al., 2011). Many genes participating in the synthesis of organic acids and
348 exoploymers were discovered, such as isopropylmalate synthase, poly-hydroxyalkanoic acid
349 synthase and polyglutamate synthase.

350 The large abundance of phosphorus (Pi) and barium (Ba) in the sediment indicated that
351 the seamount cobalt-rich ferromanganese crust region prossessed a high level of biological
352 productivity, which may directly support a rich diversity of microbes as well as benthos on the
353 seafloor. Oceanic Pi is usually enriched by marine organisms and settles into the sediment
354 (Delaney, 1998). The concentration of P_2O_5 in the sediment from station SEAM02 was 0.22%
355 higher than that from offshore sediment and continental crust (Table 1). Some genes involved
356 in the P cycle were found in the fosmid library, and the ratio of those genes to the functional
357 genes was 12.6% (22 / 174). Sedimentary Ba has been used as a proxy for the reconstruction
358 of past oceanic productivity (Schenau and De Lange, 2001). The concentration of Ba in the
359 sediment was 4.2 and 3.1 times than that in the offshore sediment and the continental crust,
360 respectively.

361 In a deep-sea environment, the electron acceptors are reduced in a sequential order based
362 on the free energy yield: O_2, NO_3^- and NO_2^-, Mn and Fe, SO_4^{2-}, and CO_2 (Wright et al., 2012).
363 The nitrogen cycle plays a key role not only in energy metabolism but also in the element
364 cycle, which affects the Mn and Fe oxidation and even mineralisation process. Ammonia
365 oxidation is a key process in marine nitrogen cycling and is executed by several microbial
366 groups, including aerobic chemoautotrophic archaea (ammonia-oxidising archaea, AOA) and
367 bacteria (ammonia-oxidising bacteria, AOB) (Beman et al., 2012; Li and Gu, 2013).
368 *Nitrosopumilus maritimus* SCM1 was the first discovered isolate of MGI Thaumarchaeota and
369 the first archaeal strain observed to undergo ammonia oxidation (Könneke et al., 2005). This
370 strain grows chemoautotrophically by oxidising ammonia aerobically and assimilating carbon
371 through the 3-hydroxypropionate/4-hydroxybutyrate pathway (Walker et al., 2010), indicating
372 that marine MGI Thaumarchaeota may be important to the nitrogen and carbon cycles in
373 ecosystem. In our study, MGI Thaumarchaeota was the most abundant archaea in the deep-sea
374 sediment (Fig. 1). Ammonia monooxygenase (AMO) is a key enzyme in the ammonia
375 oxidation process. An *amo* gene, having 98.0% amino acid sequence similarity with that from
376 *N. maritimus* SCM1, was observed in fosmid-end sequences (data not shown). Clone W4-93a
377 contained a 3-hydroxybutyryl-CoA dehydrogenase gene (ORF11), which is involved in the
378 3-hydroxypropionate/4-hydroxybutyrate cycle for autotrophic carbon fixation. Clone W4-87b,
379 a potential chemoautotrophic AOB (97.8% 16S rRNA gene identity with *Nitrosospira*

multiformis), was annotated. Although the *amo* gene was not observed in clone W4-87b, an UreD gene (ORF4), which participates in the hydrolysis of urea to ammonia, was detected. The enzyme could provide a substrate for ammonia oxidation. In addition, clone W4-93a contained putative genes for a nitrogen regulatory protein (ORF25) and a nitroreductase (ORF31). It is noteworthy that the combined microbial community and functional genes imply a nitrogen cycle was an important biogeochemical process in the deep-sea sediment from the seamount cobalt-rich ferromanganese crust region.

Our results not only support previous observations showing a relatively high abundance of Proteobacteria and MGI Thaumarchaeota in the microbial community (Liao et al., 2011) but also reveal the genomic and biological properties of the microbes in the deep-sea sediment from seamount cobalt-rich ferromanganese deposit region. In total, 78 clones containing archaeal and bacterial 16S rRNA genes were screened from 21000 clones in the metagenome library; Proteobacteria (55.8%) and MGI Thaumarchaeota (97.1%) dominated in the bacterial and archaeal communities, respectively. The continued analysis of microbial genomic data indicated that numerous genes are involved in metabolism, which is the most abundant gene group (44.0% in genes assigned to the COG classification system) (Fig. 3). Interestingly, a significant proportion of genes (8.2%) were related to DNA transport and metabolism. The plankton and microbe from the upper layers of seawater sink to the seafloor and become an energy resource there. The total DNA sinking to the seafloor was estimated to 1.26×10^7 metric tons $year^{-1}$, and up to 0.45 gigatons of extracellular DNA is present in the top 10 centimetres of deep-sea sediments (Dell'Anno and Danovaro, 2005). Recent research has found the DNA-eating ability of *Escherichia coli* during long-term survival (Finkel and Kolter, 2001; Finkel, 2006). Considering most of the planktons and microbes from the upper layers of seawater are not able to survive in the deep-sea environment, the release of their DNA into the sediment might be used as an important nutrient for indigenous microbes. Several antibiotic-synthesising genes (penicillin and streptomycin) were also detected in the metagenome library. Antibiotics can inhibit or kill some microbes, as well as benthos, and help the microbes to occupy an ecological niche. Therefore, it is essential for microbes to have alternative metabolic pathways to survive in the deep-sea environment.

Genomic divergence and HGT played important roles in the microbial adaption to the heavy metal rich and cold deep-sea environment. Previous comparative genomic studies of uncultivated marine planktonic archaea from different oceanic regions revealed significant genomic divergences, regardless of the 16S rRNA gene sequence variation (Béjà et al., 2002; Martin-Cuadrado et al., 2008). In our study, the comparative genomic analysis of clone

414 W4-93a, *N. maritimus* SCM1 and fosmid clone 74A4 suggested that considerable genome
415 divergence exists at the genus level (95.2% 16S rRNA gene sequence identity) between
416 sedimentary and planktonic lineages (Fig. 2). Although the 16S rRNA gene sequences of
417 fosmid W4-87b and *N. multiformis* ATCC 25196 showed a high identity (97.8%), their G+C
418 content (44.7% and 53.9%) and genome synteny surrounding the rRNA operon (data not
419 shown) were surprisingly different. CloneW5-51b also showed different genome synteny with
420 *Gemmatimonas aurantiaca* T-27^{T}, which is the only isolate in class Gemmatimonadetes
421 (Zhang et al., 2003). These differences may have been caused by genome evolution during
422 adaption to different habitats. Many HGT events have been observed in genomes, and the
423 HGT rate was 11.4% (23 / 201) among the seven fragments of known phylogenetic lineages
424 (W5-61a, W4-93a, W4-39b, W4-87b, W5-47b, W5-102b and W5-51b). However, 17.9% (36 /
425 201) of the predicted genes have no significant relatives, and the origin of some cannot be
426 determined due to a lack of sufficient information in the database. Taken together, the 11.4%
427 HGT rate should be most likely an underestimation. Most HGTs were from bacteria to
428 bacteria, with a few possibly were from bacteria to Thaumarchaeota, from Euryarchaeota to
429 Thaumarchaeota, from Euryarchaeota to bacteria, and even from eukarya to bacteria (Fig. S1).
430 Most of the genes acquired through HGT were involved in metabolism, including carbon and
431 energy metabolism (isopropylmalate dehydratase, pyrroloquinoline quinone biosynthesis
432 protein C, glutamate synthase, cytochrome c family protein, MIP family channel protein and
433 alcohol dehydrogenase) and nitrogen metabolism (nitrogen regulatory protein and urease
434 accessory protein UreD).
435 In conclusion, element concentrations in the sediment from the seamount cobalt-rich
436 ferromanganese crust region are different from those in other marine or terrestrial
437 environments. The large abundances of heavy metals (Mn, Ni, Co, Cu), P and Ba in the
438 sediment from station SEAM02 implied a unique microbial community with high biodiversity.
439 Microbes inhabiting the cobalt-rich ferromanganese crust region not only adapt to high
440 amounts of heavy metal but also might participate in the biomineralization process, as
441 observed at the gene level. Alternative metabolic pathways and a variety of stress genes are
442 essential for microbial survival in the deep-sea environment. Genomic divergence and HGT
443 may played important roles in the microbial adaption to the deep-sea environment. Some
444 microbes, which come from the upper seawater, might obtain a series of new features and
445 adapt to this harsh environment via high frequency HGT events. The information gathered via
446 the rRNA-gene based PCR screening method provided insight only into the genomic regions
447 directly adjacent to rRNA operons. However, this is the first metagenomic study of deep-sea

sediment from the cobalt-rich ferromanganese crust region, giving us some insights into the genetic and functional information about uncultured microorganisms in the cobalt-rich ferromanganese crust region.

References

Adachi O, Ano Y, Toyama H, Matsushita K (2007). Biooxidation with PQQ- and FAD-dependent dehydrogenases. *Modern Biooxidation*. Wiley-VCH Verlag GmbH & Co. KGaA. pp 1-41.

Béjà O, Koonin EV, Aravind L, Taylor LT, Seitz H, Stein JL *et al* (2002). Comparative genomic analysis of archaeal genotypic variants in a single population and in two different oceanic provinces. *Applied and Environmental Microbiology* **68:** 335-345.

Beman JM, Popp BN, Alford SE (2012). *Quantification of ammonia oxidation rates and ammonia-oxidizing archaea and bacteria at high resolution in the Gulf of California and eastern tropical North Pacific Ocean*, vol. 57. American Society of Limnology and Oceanography: Waco, TX, ETATS-UNIS.

Candela T, Fouet A (2006). Poly-gamma-glutamate in bacteria. *Molecular Microbiology* **60:** 1091-1098.

Clark MR, Rowden AA, Schlacher T, Williams A, Consalvey M, Stocks KI *et al* (2010). The ecology of seamounts: structure, function, and human impacts. *Ann Rev Mar Sci* **2:** 253-278.

Craig JD, Andrews JE, Meylan MA (1982). Ferromanganese deposits in the Hawaiian Archipelago. *Marine Geology* **45:** 127-157.

Delaney ML (1998). Phosphorus accumulation in marine sediments and the oceanic phosphorus cycle. *Global Biogeochemical Cycles* **12:** 563-572.

Dell'Anno A, Danovaro R (2005). Extracellular DNA plays a key role in deep-sea ecosystem functioning. *Science* **309:** 2179.

DeLong EF (1992). Archaea in coastal marine environments. *Proceedings of the National Academy of Sciences of the United States of America* **89:** 5685-5689.

Duffy EJ (2008). *"Seamount". CenSeam: a Global Census of Marine Life on Seamounts (Content Partner) and National Oceanic and Atmospheric Administration (Content source)*. Environmental Information Coalition, National Council for Science and the Environment: Washington D.C.

Ehrhardt CJ, Haymon RM, Lamontagne MG, Holden PA (2007). Evidence for hydrothermal Archaea within the basaltic flanks of the East Pacific Rise. *Environmental Microbiology* **9:** 900-912.

Emerson D, Moyer CL (2010). Microbiology of seamounts: common patterns observed in community structure. *Oceanography* **23:** 148-163.

Finkel SE, Kolter R (2001). DNA as a nutrient: novel role for bacterial competence gene homologs. *Journal of Bacteriology* **183:** 6288-6293.

Finkel SE (2006). Long-term survival during stationary phase: evolution and the GASP phenotype. *Nat Rev Microbiol* **4:** 113-120.

Fu Y, Peng J, Qu W, Hu R, Shi X, Du A (2005). Os isotopic compositions of a cobalt-rich ferromanganese crust profile in Central Pacific. *ChinSciBull* **50:** 2106-2112.

Fuyuan Z, Weiyan Z, Kechao ZHU, Shuitu GAO, Haisheng Z, Xiaoyu Z *et al* (2008). Distribution characteristics of cobalt-rich ferromanganese crust resources on submarine seamounts in the Western Pacific. *Acta Geologica Sinica - English Edition* **82:** 796-803.

Gadd GM (2007). Geomycology: biogeochemical transformations of rocks, minerals, metals and radionuclides by fungi, bioweathering and bioremediation. *Mycological Research* **111:** 3-49.

Garcia-Martinez J, Martinez-Murcia A, Anton AI, Rodriguez-Valera F (1996). Comparison of the small 16S to 23S intergenic spacer region (ISR) of the rRNA operons of some Escherichia coli strains of the ECOR collection and E. coli K-12. *Journal of Bacteriology* **178:** 6374-6377.

Ghiorse WC (1984). Biology of iron- and manganese-depositing bacteria. *Annu Rev Microbiol* **38:** 515-550.

Gilbert J, Blainey PC, Mosier AC, Potanina A, Francis CA, Quake SR (2011). Genome of a low-salinity ammonia-oxidizing archaeon determined by single-cell and metagenomic analysis. *PLoS ONE* **6:** e16626.

Gillan DC, Danis B (2007). The archaebacterial communities in Antarctic bathypelagic sediments. *Deep Sea Research Part II: Topical Studies in Oceanography* **54:** 1682-1690.

Guibaud G, van Hullebusch E, Bordas F, d'Abzac P, Joussein E (2009). Sorption of Cd(II) and Pb(II) by exopolymeric substances (EPS) extracted from activated sludges and pure bacterial strains: Modeling of the metal/ligand ratio effect and role of the mineral fraction. *Bioresource Technology* **100:** 2959-2968.

Heijs SK, Haese RR, Wielen PWJJ, Forney LJ, Elsas JD (2007). Use of 16S rRNA gene based clone libraries to assess microbial communities potentially involved in anaerobic methane oxidation in a Mediterranean cold seep. *Microbial Ecology* **53:** 384-398.

Hillier JK, Watts AB (2007). Global distribution of seamounts from ship-track bathymetry data. *Geophysical Research Letters* **34:** L13304.

Ito M, Tsunekawa M, Yamaguchi E, Sekimura K, Kashiwaya K, Hori K *et al* (2008). Estimation of degree of liberation in a coarse crushed product of cobalt-rich ferromanganese crust/nodules and its gravity separation. *International Journal of Mineral Processing* **87:** 100-105.

Iyer SD, M. MC, Das P, Kalangutkar NG (2012). Seamounts - characteristics, formation, mineral deposits and biodiversity. *Geologica Acta* **10:** 295-308.

Jiang X, Xu X, Huo Y, Wu Y, Zhu X, Zhang X *et al* (2012). Identification and characterization of novel esterases from a deep-sea sediment metagenome. *Archives of Microbiology* **194:** 207-214.

Könneke M, Bernhard AE, de la Torre JR, Walker CB, Waterbury JB, Stahl DA (2005). Isolation of an autotrophic ammonia-oxidizing marine archaeon. *Nature* **437:** 543-546.

Kanehisa M, Goto S (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research* **28:** 27-30.

Kato S, Kobayashi C, Kakegawa T, Yamagishi A (2009a). Microbial communities in iron-silica-rich microbial mats at deep-sea hydrothermal fields of the Southern Mariana Trough. *Environmental Microbiology* **11:** 2094-2111.

Kato S, Yanagawa K, Sunamura M, Takano Y, Ishibashi J-i, Kakegawa T *et al* (2009b). Abundance of Zetaproteobacteria within crustal fluids in back-arc hydrothermal fields of the Southern Mariana Trough. *Environmental Microbiology* **11:** 3210-3222.

Koschinsky A, Hein JR (2003). Uptake of elements from seawater by ferromanganese crusts: solid-phase associations and seawater speciation. *Marine Geology* **198:** 331-351.

Li M, Gu J-D (2013). Community structure and transcript responses of anammox bacteria, AOA, and AOB in mangrove sediment microcosms amended with ammonium and nitrite. *Applied Microbiology and Biotechnology***:** 1-16.

Li Y, Li F, Zhang X, Qin S, Zeng Z, Dang H *et al* (2008). Vertical distribution of bacterial and archaeal communities along discrete layers of a deep-sea cold sediment sample at the East Pacific Rise (~13°N). *Extremophiles* **12:** 573-585.

Liao L, Xu X-W, Jiang X-W, Wang C-S, Zhang D-S, Ni J-Y *et al* (2011). Microbial diversity in deep-sea sediment from the cobalt-rich crust deposit region in the Pacific Ocean. *FEMS Microbiology Ecology* **78:** 565-585.

Lopez-Diaz I, Clarke S, Mandelstam J (1986). spoIID Operon of Bacillus subtilis: cloning and sequence. *Journal of General Microbiology* **132:** 341-354.

Losekann T, Knittel K, Nadalig T, Fuchs B, Niemann H, Boetius A *et al* (2007). Diversity and abundance of aerobic and anaerobic methane oxidizers at the Haakon Mosby Mud Volcano, Barents Sea. *Applied and Environmental Microbiology* **73:** 3348-3362.

Martín-Cuadrado A-B, López-García P, Alba J-C, Moreira D, Monticelli L, Strittmatter A *et al* (2007). Metagenomics of the deep Mediterranean, a warm bathypelagic habitat. *PLoS ONE* **2:** e914.

Martin-Cuadrado A-B, Rodriguez-Valera F, Moreira D, Alba JC, Ivars-Martínez E, Henn MR *et al* (2008). Hindsight in the relative abundance, metabolic potential and genome dynamics of uncultivated marine archaea from comparative metagenomic analyses of bathypelagic plankton of different oceanic regions. *The ISME Journal* **2:** 865-886.

Menard HW (1964). *Marine Geology of the Pacific*. McGraw-Hill: New York.

Miroshnichenko M, Hippe H, Stackebrandt E, Kostrikina N, Chernyh N, Jeanthon C *et al* (2001). Isolation and characterization of Thermococcus sibiricus sp. nov. from a Western Siberia high-temperature oil reservoir. *Extremophiles* **5:** 85-91.

Muiños SB, Hein JR, Frank M, Monteiro JH, Gaspar L, Conrad T *et al* (2013). Deep-sea Fe-Mn crusts from the Northeast Atlantic Ocean: composition and resource considerations. *Marine Georesources & Geotechnology* **31:** 40-70.

Nies DH (1992). Resistance to cadmium, cobalt, zinc, and nickel in microbes. *Plasmid* **27:** 17-28.

Norton JM, Klotz MG, Stein LY, Arp DJ, Bottomley PJ, Chain PSG *et al* (2008). Complete genome sequence of Nitrosospira multiformis, an ammonia-oxidizing bacterium from the soil environment. *Applied and Environmental Microbiology* **74:** 3559-3572.

Rowden AA, Dower JF, Schlacher TA, Consalvey M, Clark MR (2010). Paradigms in seamount ecology: fact, fiction and future. *Marine Ecology* **31:** 226-241.

Santelli CM, Orcutt BN, Banning E, Bach W, Moyer CL, Sogin ML *et al* (2008). Abundance and diversity of microbial life in ocean crust. *Nature* **453:** 653-656.

Santelli CM, Webb SM, Dohnalkova AC, Hansel CM (2011). Diversity of Mn oxides produced by Mn(II)-oxidizing fungi. *Geochimica et Cosmochimica Acta* **75:** 2762-2776.

Schauer R, Bienhold C, Ramette A, Harder J (2009). Bacterial diversity and biogeography in deep-sea surface sediments of the South Atlantic Ocean. *The ISME Journal* **4:** 159-170.

Schenau SJ, De Lange GJ (2001). Phosphorus regeneration vs. burial in sediments of the Arabian Sea. *Marine Chemistry* **75:** 201-217.

Schlacher TA, Rowden AA, Dower JF, Consalvey M (2010). Seamount science scales undersea mountains: new research and outlook. *Marine Ecology* **31:** 1-13.

Staudigel H, Koppers AAP, Plank TA, Hanan BB (2010). Seamounts in the subduction factory. *Oceanography* **23:** 176-181.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011). MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Molecular Biology and Evolution* **28:** 2731-2739.

Tatusov RL, Galperin MY, Natale DA, Koonin EV (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Research* **28:** 33-36.

Tebo BM, Bargar JR, Clement BG, Dick GJ, Murray KJ, Parker D *et al* (2004). Biogenic manganese oxides: properties and mechanisms of formation. *Annual Review of Earth and Planetary Sciences* **32:** 287-328.

Verlaan PA (1992). Benthic recruitment and manganese crust formation on seamounts. *Marine Biology* **113:** 171-174.

Walker CB, de la Torre JR, Klotz MG, Urakawa H, Pinel N, Arp DJ *et al* (2010). Nitrosopumilus maritimus genome reveals unique mechanisms for nitrification and autotrophy in globally distributed marine crenarchaea. *Proceedings of the National Academy of Sciences of the United States of America* **107:** 8818-8823.

Wang X-H, Schlossmacher U, Natalio F, Schroder HC, Wolf SE, Tremel W *et al* (2009). Evidence for biogenic processes during formation of ferromanganese crusts from the Pacific Ocean: implications of biologically induced mineralization. *Micron* **40:** 526-535.

Wang X, Müller WEG (2009). Marine biominerals: perspectives and challenges for polymetallic nodules and crusts. *Trends in Biotechnology* **27:** 375-383.

Wang X, Wiens M, Schröder H, Schloßmacher U, Müller WG (2011). Molecular biomineralization: toward an understanding of the biogenic origin of polymetallic nodules, seamount crusts, and hydrothermal vents. In: Müller WEG (ed). *Molecular Biomineralization.*

Springer Berlin Heidelberg. pp 77-110.

Wedepohl KH (1995). The composition of the continental crust. *Geochimica et Cosmochimica Acta* **59:** 1217-1232.

Wessel P, Sandwell DT, S.-S. K (2010). The global seamount census. *Oceanography* **23:** 115-129.

Wright JJ, Konwar KM, Hallam SJ (2012). Microbial ecology of expanding oxygen minimum zones. *Nat Rev Microbiol* **10:** 381-394.

Wu Y-H, Liao L, Wang C-S, Ma W-L, Meng F-X, Wu M *et al* (2013). A comparison of microbial communities in deep-sea polymetallic nodules and the surrounding sediments in the Pacific Ocean. *Deep Sea Research Part I: Oceanographic Research Papers* **79:** 40-49.

Xu M, Wang P, Wang F, Xiao X (2005). Microbial diversity at a deep-sea station of the Pacific nodule province. *Biodiversity and Conservation* **14:** 3363-3380.

Yesson C, Clark MR, Taylor ML, Rogers AD (2011). The global distribution of seamounts based on 30 arc seconds bathymetry data. *Deep Sea Research Part I: Oceanographic Research Papers* **58:** 442-453.

Zhang H, Sekiguchi Y, Hanada S, Hugenholtz P, Kim H, Kamagata Y *et al* (2003). Gemmatimonas aurantiaca gen. nov., sp. nov., a Gram-negative, aerobic, polyphosphate-accumulating micro-organism, the first cultured representative of the new bacterial phylum Gemmatimonadetes phyl. nov. *International Journal of Systematic and Evolutionary Microbiology* **53:** 1155-1163.

Zhao Q (1988). *Ocean geochemistry*. the Geological Publishing House: Beijing.

Zhu W, Lomsadze A, Borodovsky M (2010). Ab initio gene identification in metagenomic sequences. *Nucleic Acids Research* **38:** e132-e132.

699 Table 1. Geochemical properties of the sediment sample SEAM02, offshore sediment and
700 continental crust

	SEAM02	Offshore Sediment*	Continental Crust†
SiO_2 /%	54.07	54.43	61.50
Al_2O_3 /%	8.35	12.03	15.10
Fe_2O_3 /%	4.14	4.59	6.28
CaO /%	0.92	10.05	5.50
MgO /%	2.58	1.84	3.70
K_2O /%	2.10	1.98	2.40
Na_2O /%	7.18	2.24	3.20
TiO_2 /%	0.48	0.57	0.68
MnO /%	0.33	0.12	0.10
P_2O_5/%	0.22	0.12	0.18
Ni /μg g-1	74	27	56
Co /μg g-1	52	14	24
V /μg g-1	94	-	98
Cr /μg g-1	57	43	126
Cu /μg g-1	134	20	25
Zn /μg g-1	83	72	65
Cd /μg g-1	0.09	-	-
Pb /μg g-1	35	28	14.8
Mo /μg g-1	4.4	-	-
Ba /μg g-1	1821	431	584

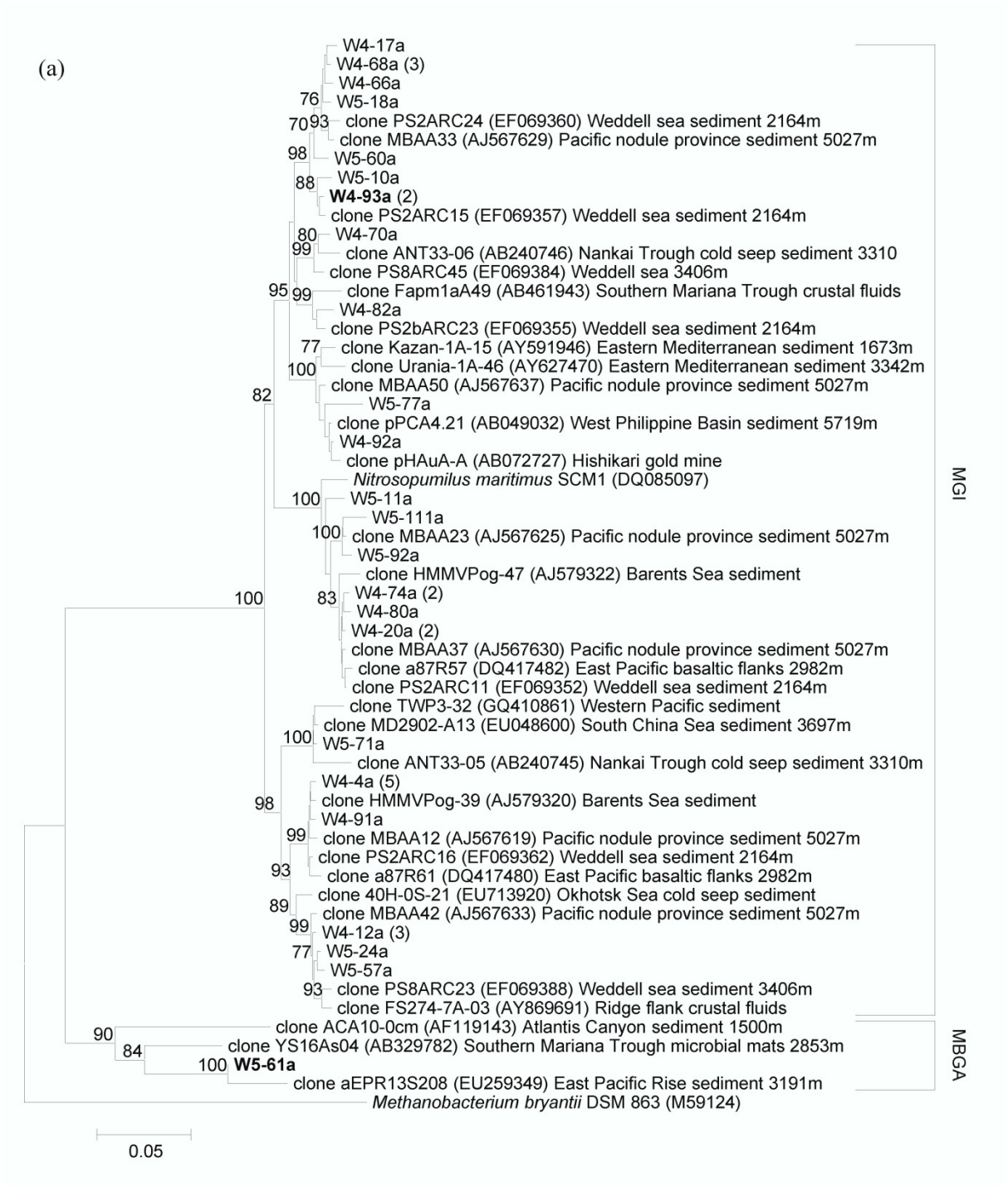
701 *Data of the East China Sea from Zhao, 1988;
702 †Data from Wedepohl, 1995.
703

Fig. 1. Neighbour-joining trees of archaeal (a) and bacterial (b) 16S rRNA genes amplified from the metagenomic library SEAM02. Numbers in parentheses represent the total number of clones with the same 16S rRNA sequence. The environments where the relatives were obtained from are given after their NCBI accession numbers. Numbers at nodes correspond to bootstrap values based on 500 replicates, and the values less than 70% were omitted. Bar, 0.05 substitutions per nucleotide position.

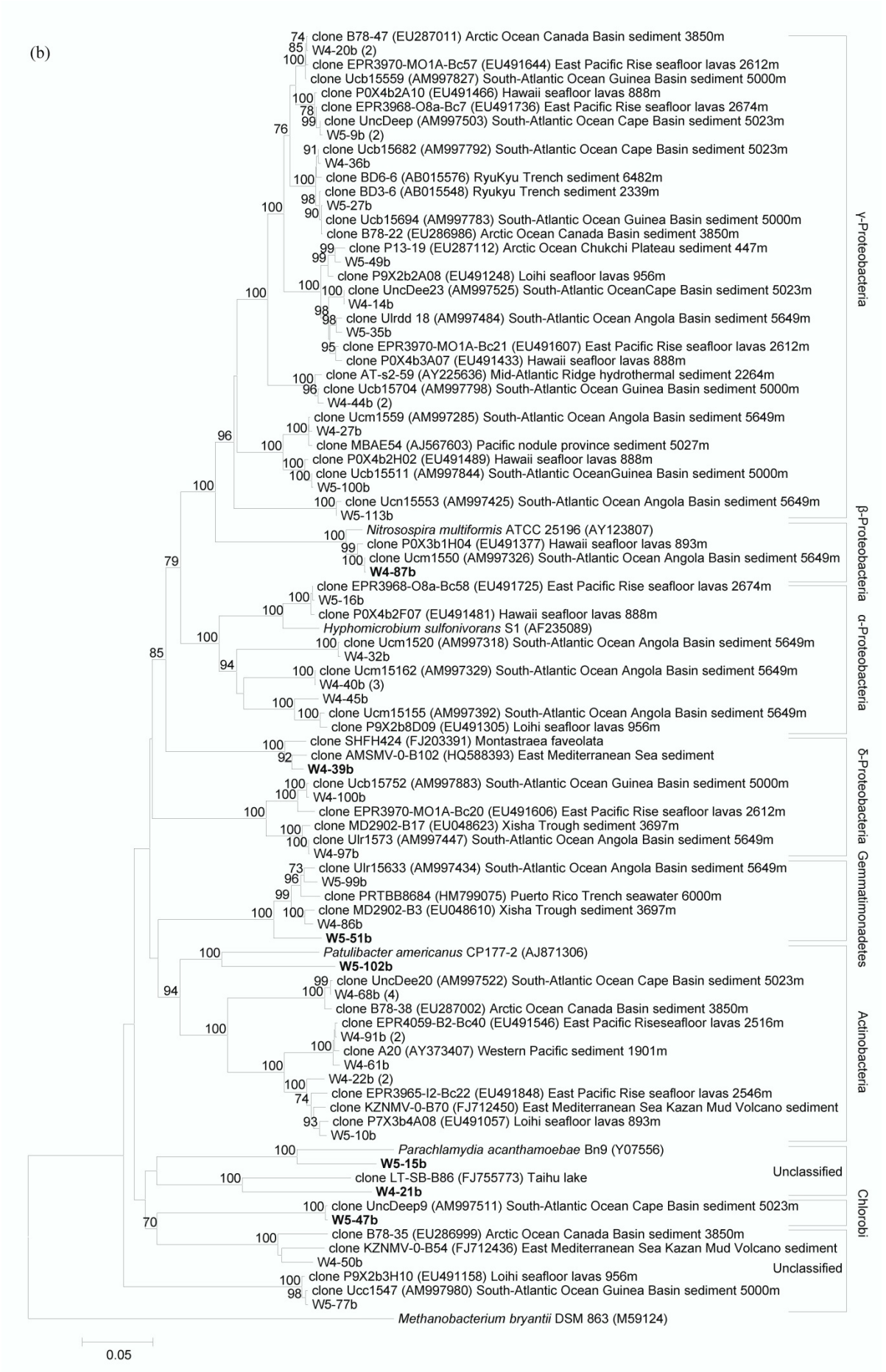
Fig. 2. Gene maps of fosmid clone W4-93a, *Nitrosopumilus maritimus* SCM1 and fosmid 74A4. The rRNA genes were used as an alignment point. ORFs: 1, ATP-dependent DNA ligase; 3, hypothetical protein; 4, fructose-1,6-bisphosphatase; 5, translation elongation factor EF-1 alpha; 6, ribosomal protein S10; 7, RNA polymerase Rbp10; 8, C2H2 Zn finger protein; 9, hypothetical protein; 11, rossmann fold nucleotide-binding protein; 12, 3-hydroxybutyryl-CoA dehydrogenase; 13, hypothetical protein; 14, HIT superfamily hydrolase; 15, DnaJ class molecular chaperone; 18, TPR repeat-containing protein; 23, glutamate-1-semialdehyde aminotransferase; 30, hypothetical protein; 34, peptide methionine sulfoxide reductase; 36, alpha/beta hydrolase; 37, AbrB family transcription regulator; 38, hypothetical protein; 39, poly(R)-hydroxyalkanoic acid synthase subunit PhaC; 40, hypothetical protein; 41, hypothetical protein; 44, transcription factor TFIIB cyclin-related protein; 46, TPR repeat-containing protein; 48, biotin--acetyl-CoA-carboxylase ligase.

Fig. 3. Functional classification of genes of the nine fosmids according to COG classification system. Blue, the COG corresponding to the “cellular processes and signalling”; green - “information storage and processing”; red - “metabolism”; orange - “poorly characterised”.

727 Fig. 1.

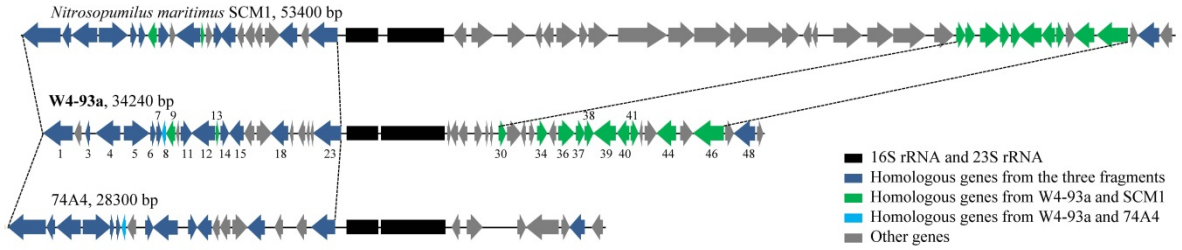

(a)
W4-17a
W4-68a (3)
W4-66a
W5-18a
clone PS2ARC24 (EF069360) Weddell sea sediment 2164m
clone MBAA33 (AJ567629) Pacific nodule province sediment 5027m
W5-60a
W5-10a
W4-93a (2)
clone PS2ARC15 (EF069357) Weddell sea sediment 2164m
W4-70a
clone ANT33-06 (AB240746) Nankai Trough cold seep sediment 3310
clone PS8ARC45 (EF069384) Weddell sea 3406m
clone Fapm1aA49 (AB461943) Southern Mariana Trough crustal fluids
W4-82a
clone PS2bARC23 (EF069355) Weddell sea sediment 2164m
clone Kazan-1A-15 (AY591946) Eastern Mediterranean sediment 1673m
clone Urania-1A-46 (AY627470) Eastern Mediterranean sediment 3342m
clone MBAA50 (AJ567637) Pacific nodule province sediment 5027m
W5-77a
clone pPCA4.21 (AB049032) West Philippine Basin sediment 5719m
W4-92a
clone pHAuA-A (AB072727) Hishikari gold mine
Nitrosopumilus maritimus SCM1 (DQ085097)
W5-11a
W5-111a
clone MBAA23 (AJ567625) Pacific nodule province sediment 5027m
W5-92a
clone HMMVPog-47 (AJ579322) Barents Sea sediment
W4-74a (2)
W4-80a
W4-20a (2)
clone MBAA37 (AJ567630) Pacific nodule province sediment 5027m
clone a87R57 (DQ417482) East Pacific basaltic flanks 2982m
clone PS2ARC11 (EF069352) Weddell sea sediment 2164m
clone TWP3-32 (GQ410861) Western Pacific sediment
clone MD2902-A13 (EU048600) South China Sea sediment 3697m
W5-71a
clone ANT33-05 (AB240745) Nankai Trough cold seep sediment 3310m
W4-4a (5)
clone HMMVPog-39 (AJ579320) Barents Sea sediment
W4-91a
clone MBAA12 (AJ567619) Pacific nodule province sediment 5027m
clone PS2ARC16 (EF069362) Weddell sea sediment 2164m
clone a87R61 (DQ417480) East Pacific basaltic flanks 2982m
clone 40H-0S-21 (EU713920) Okhotsk Sea cold seep sediment
clone MBAA42 (AJ567633) Pacific nodule province sediment 5027m
W4-12a (3)
W5-24a
W5-57a
clone PS8ARC23 (EF069388) Weddell sea sediment 3406m
clone FS274-7A-03 (AY869691) Ridge flank crustal fluids
clone ACA10-0cm (AF119143) Atlantis Canyon sediment 1500m
clone YS16As04 (AB329782) Southern Mariana Trough microbial mats 2853m
W5-61a
clone aEPR13S208 (EU259349) East Pacific Rise sediment 3191m
Methanobacterium bryantii DSM 863 (M59124)
MGI
MBGA
0.05

728


(b)
clone B78-47 (EU287011) Arctic Ocean Canada Basin sediment 3850m
W4-20b (2)
clone EPR3970-MO1A-Bc57 (EU491644) East Pacific Rise seafloor lavas 2612m
clone Ucb15559 (AM997827) South-Atlantic Ocean Guinea Basin sediment 5000m
clone P0X4b2A10 (EU491466) Hawaii seafloor lavas 888m
clone EPR3968-O8a-Bc7 (EU491736) East Pacific Rise seafloor lavas 2674m
clone UncDeep (AM997503) South-Atlantic Ocean Cape Basin sediment 5023m
W5-9b (2)
clone Ucb15682 (AM997792) South-Atlantic Ocean Cape Basin sediment 5023m
W4-36b
clone BD6-6 (AB015576) RyuKyu Trench sediment 6482m
clone BD3-6 (AB015548) Ryukyu Trench sediment 2339m
W5-27b
clone Ucb15694 (AM997783) South-Atlantic Ocean Guinea Basin sediment 5000m
clone B78-22 (EU286986) Arctic Ocean Canada Basin sediment 3850m
clone P13-19 (EU287112) Arctic Ocean Chukchi Plateau sediment 447m
W5-49b
clone P9X2b2A08 (EU491248) Loihi seafloor lavas 956m
clone UncDee23 (AM997525) South-Atlantic OceanCape Basin sediment 5023m
W4-14b
clone Ulrdd 18 (AM997484) South-Atlantic Ocean Angola Basin sediment 5649m
W5-35b
clone EPR3970-MO1A-Bc21 (EU491607) East Pacific Rise seafloor lavas 2612m
clone P0X4b3A07 (EU491433) Hawaii seafloor lavas 888m
clone AT-s2-59 (AY225636) Mid-Atlantic Ridge hydrothermal sediment 2264m
clone Ucb15704 (AM997798) South-Atlantic Ocean Guinea Basin sediment 5000m
W4-44b (2)
clone Ucm1559 (AM997285) South-Atlantic Ocean Angola Basin sediment 5649m
W4-27b
clone MBAE54 (AJ567603) Pacific nodule province sediment 5027m
clone P0X4b2H02 (EU491489) Hawaii seafloor lavas 888m
clone Ucb15511 (AM997844) South-Atlantic OceanGuinea Basin sediment 5000m
W5-100b
clone Ucn15553 (AM997425) South-Atlantic Ocean Angola Basin sediment 5649m
W5-113b
Nitrosospira multiformis ATCC 25196 (AY123807)
clone P0X3b1H04 (EU491377) Hawaii seafloor lavas 893m
clone Ucm1550 (AM997326) South-Atlantic Ocean Angola Basin sediment 5649m
W4-87b
clone EPR3968-O8a-Bc58 (EU491725) East Pacific Rise seafloor lavas 2674m
W5-16b
clone P0X4b2F07 (EU491481) Hawaii seafloor lavas 888m
Hyphomicrobium sulfonivorans S1 (AF235089)
clone Ucm1520 (AM997318) South-Atlantic Ocean Angola Basin sediment 5649m
W4-32b
clone Ucm15162 (AM997329) South-Atlantic Ocean Angola Basin sediment 5649m
W4-40b (3)
W4-45b
clone Ucm15155 (AM997392) South-Atlantic Ocean Angola Basin sediment 5649m
clone P9X2b8D09 (EU491305) Loihi seafloor lavas 956m
clone SHFH424 (FJ203391) Montastraea faveolata
clone AMSMV-0-B102 (HQ588393) East Mediterranean Sea sediment
W4-39b
clone Ucb15752 (AM997883) South-Atlantic Ocean Guinea Basin sediment 5000m
W4-100b
clone EPR3970-MO1A-Bc20 (EU491606) East Pacific Rise seafloor lavas 2612m
clone MD2902-B17 (EU048623) Xisha Trough sediment 3697m
clone Ulr1573 (AM997447) South-Atlantic Ocean Angola Basin sediment 5649m
W4-97b
clone Ulr15633 (AM997434) South-Atlantic Ocean Angola Basin sediment 5649m
W5-99b
clone PRTBB8684 (HM799075) Puerto Rico Trench seawater 6000m
clone MD2902-B3 (EU048610) Xisha Trough sediment 3697m
W4-86b
W5-51b
Patulibacter americanus CP177-2 (AJ871306)
W5-102b
clone UncDee20 (AM997522) South-Atlantic Ocean Cape Basin sediment 5023m
W4-68b (4)
clone B78-38 (EU287002) Arctic Ocean Canada Basin sediment 3850m
clone EPR4059-B2-Bc40 (EU491546) East Pacific Riseseafloor lavas 2516m
W4-91b (2)
clone A20 (AY373407) Western Pacific sediment 1901m
W4-61b
W4-22b (2)
clone EPR3965-I2-Bc22 (EU491848) East Pacific Rise seafloor lavas 2546m
clone KZNMV-0-B70 (FJ712450) East Mediterranean Sea Kazan Mud Volcano sediment
clone P7X3b4A08 (EU491057) Loihi seafloor lavas 893m
W5-10b
Parachlamydia acanthamoebae Bn9 (Y07556)
W5-15b
clone LT-SB-B86 (FJ755773) Taihu lake
W4-21b
clone UncDeep9 (AM997511) South-Atlantic Ocean Cape Basin sediment 5023m
W5-47b
clone B78-35 (EU286999) Arctic Ocean Canada Basin sediment 3850m
clone KZNMV-0-B54 (FJ712436) East Mediterranean Sea Kazan Mud Volcano sediment
W4-50b
clone P9X2b3H10 (EU491158) Loihi seafloor lavas 956m
clone Ucc1547 (AM997980) South-Atlantic Ocean Guinea Basin sediment 5000m
W5-77b
Methanobacterium bryantii DSM 863 (M59124)
0.05
γ-Proteobacteria
β-Proteobacteria
α-Proteobacteria
δ-Proteobacteria
Gemmatimonadetes
Actinobacteria
Unclassified
Chlorobi
Unclassified

729

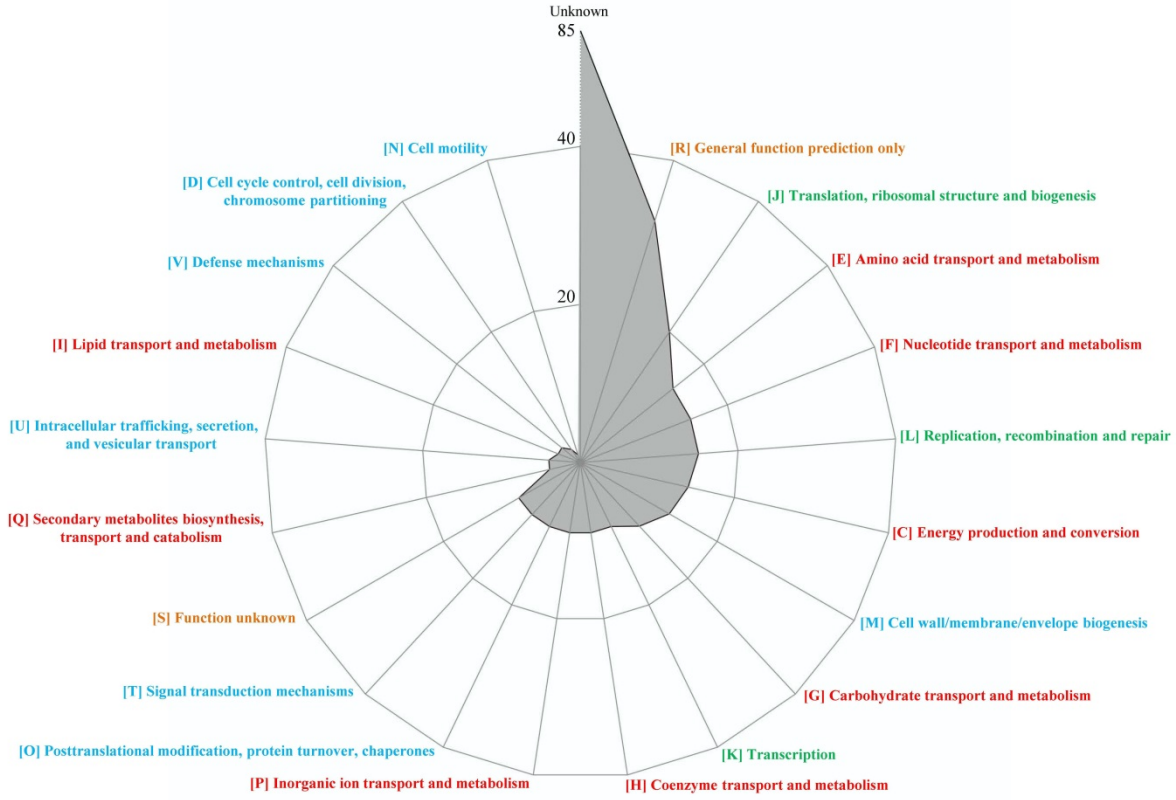
730 Fig. 2.



731

732

733 Fig. 3.



734